**Analisi di dati sulle forniture elettriche: campionamento stratificato e procedure di clustering su utenze trifase e monofase**

Corso di Laurea Magistrale in Ingegneria Elettrica

Candidato: *Daniele Bonansinga*, Relatore: *Gianfranco Chicco*

Marzo 2022, Anno Accademico 2021/2022

**Abstract- Il lavoro che viene presentato si concentra sull’acquisizione, elaborazione e gestione di un gran numero di dati raccolti dalla installazione degli smart meter in un sistema di distribuzione. L’impossibilità di analizzare tutte le informazioni utili porta a studiare nuove metodologie per determinare il corretto numero di dati da analizzare cercando di abbassare il più possibile la soglia di errore nella stima finale. La teoria del campionamento stratificato è essenziale per ridurre l’onere computazionale dell’analisi dei dati e far sì che ogni macro-classe sia rappresentata con un numero appropriato di utenti. Grazie a diverse procedure di clustering si possono creare nuovi gruppi di clienti per similarità. L’analisi permette lo sviluppo di nuove applicazioni, come una nuova classificazione dei clienti; infatti, si può abbandonare il concetto di classificazione dei clienti basata sui parametri commerciali, poiché i diversi raggruppamenti vengono creati a partire dallo studio delle caratteristiche di consumo di energia di ogni utente, determinate ad esempio attraverso lo studio delle curve di carico.**

**Ulteriori sviluppi di questo studio potranno ammodernare alcune applicazioni specifiche come la previsione del carico e l’analisi di nuovi carichi aggregati anche di tipo attivo.**

***Parole chiave- Campionamento stratificato, procedure di clustering, k-means, clustering gerarchico, smart meter.***

I. INTRODUZIONE

La progressiva installazione dei contatori di ultima generazione, sugli utenti finali, sta mettendo a disposizione un gigantesco numero di dati con risoluzioni anche inferiori al quarto d’ora. La gestione di questo enorme flusso di dati non è semplice, per questo motivo si stanno sviluppando nuove interfacce di dati e applicazioni che estraggono i dati in serie temporali. Per capire la reale potenzialità degli smart meter nella prima parte della tesi verranno confrontate le due diverse generazioni di smart meter 1G e 2G.

Una volta appreso l’enorme quantitativo di dati utili che può raccogliere il contatore 2G, si sono presi in esame i dati sul consumo e sulle curve di carico degli utenti monofase a uso residenziale e trifase a uso non residenziale degli utenti di una rete di distribuzione dell’anno 2020. Naturalmente tutte le forniture sono state analizzate in forma anonima.

Una volta ridotto l’onere dei dati da più di un milione a qualche migliaio grazie al metodo del campionamento stratificato, sono state utilizzate due procedure di clustering, clustering gerarchico e l’algoritmo k-means, per la creazione di gruppi e sottogruppi utili per dare al distributore e al cliente finale una consapevolezza dei consumi che variano nel tempo. Da questa analisi si determinano nuove classi di utenti partendo dalle curve di carico rappresentative mensili e annuali.

II. Campionamento stratificato

I clienti elettrici sono normalmente suddivisi in macro-classi dipendenti dal tipo d’uso dell’energia elettrica e dalla loro potenza contrattuale. Per determinare delle micro-classi per similarità delle loro curve di carico, il metodo più rigoroso sarebbe analizzare ogni curva di carico di ogni giorno del 2020 di tutti gli utenti forniti. Visto che l’insieme dei clienti è molto numeroso, tale approccio risulta poco pratico e molto costoso. Siamo costretti ad utilizzare alcune tecniche statistiche significative per determinare il miglior campionamento di ogni classe di clienti. Nel caso preso in esame si è scelto di utilizzare la teoria del campionamento stratificato. Una volta determinato il numero suggerito di campioni che rappresentano il numero di utenti totali, con un intervallo di confidenza accettabile, si possono analizzare i dati delle forniture campione. Esse sono numericamente distribuite come da output del campionamento stratificato, ricoprendo tutte le macro-classi con una stima di errore accettabile.

L’applicazione del metodo è stata effettuata sui dati disponibili, per ogni utente di Bassa Tensione nell’anno 2020, della potenza contrattuale, del consumo energetico e delle curve di carico. Il metodo del campionamento stratificato è stato applicato a due tipologie di clienti differenti: utenti trifase a uso non residenziale e utenti monofase a uso residenziale. Il primo passo è stato suddividere gli utenti in macro-classi secondo la tipologia d’uso e secondo la loro potenza contrattuale. Per gli utenti trifase sono state trovate dieci macro-classi ORT (la sigla ORT sta per gli utenti con tipologia d’uso ordinaria trifase) e dieci macro-classi SAT (dove SAT sta per utenti con tipologia d’uso straordinaria trifase). Lo stesso ragionamento è stato utilizzato per suddividere gli utenti monofase. L’analisi per entrambe le macro-classi è stata condotta suddividendo due casi, ovvero considerando due variabili discriminanti diverse. Nel primo caso è stato analizzato il campionamento stratificato usando come variabile discriminante la potenza contrattuale. Come secondo caso è stata utilizzata come variabile discriminante l’energia consumata. Una volta calcolato il numero di ripartizioni per ogni classe, output del campionamento stratificato, è stata studiata la stima dell’intervallo di confidenza. Questo risultato è stato rappresentato graficamente, cosicché si può osservare la stima dell’intervallo di confidenza per diversi numeri totali di utenti. Questo ragionamento è stato fatto per tutti i mesi dell’anno.



2020.*Grafico 1: Variazione dell’intervallo di confidenza con caratteristica variabile l’energia consumata per le forniture trifase a uso non residenziale del mese di marzo 2020.*

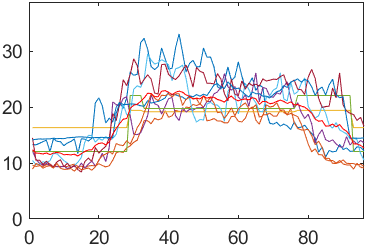
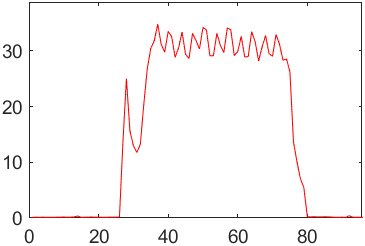
Avendo trovato l’output del campionamento stratificato, è stato deciso quale stima dell’intervallo di confidenza prendere. Per questo, occorre scegliere il numero di utenti totale da analizzare e quindi capire la loro distribuzione all’interno delle differenti classi. Si è scelto di prendere 2500 utenti, distribuiti come nel caso dell’energia consumata, accettando così una stima dell’intervallo di confidenza che va dall’1.5%, nel caso migliore e il 6% nel caso peggiore. Questa scelta è stata fatta poiché i dati sono distribuiti più uniformemente all’interno delle varie classi rispetto al caso della potenza contrattuale.

III. CREAZIONE DEI CLUSTER ANALIZZANDO LE CURVE DI CARICO

Per la creazione dei vari cluster, bisogna andare ad analizzare le curve di carico dei clienti secondo la ripartizione scelta dal campionamento stratificato.

Bisogna creare delle curve di carico rappresentative di ogni fornitura sia mensili che annuali. Una volta create si può procedere con la creazione delle micro-classi tramite diverse procedure di clustering.

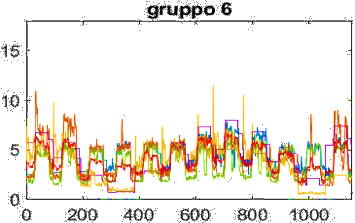
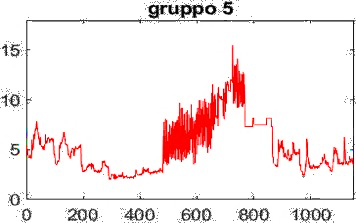
I diversi cluster sono stati creati utilizzando l’algoritmo di clustering gerarchico. Utilizzando questo metodo si possono vedere i vari raggruppamenti di carichi per similarità e i principali outlier. Nel grafico 2 e nel grafico 3 possiamo vedere degli esempi di outlier e di alcune micro-classi create grazie alla procedura di clustering (l’asse orizzontale riporta i quarti d’ora).



quarto d’ora

quarto d’ora

*Grafico 2: esempi di cluster creati dall’algoritmo gerarchico, macro-classe ort9 (mese di gennaio), curve di carico rappresentative mensili.*



quarto d’ora

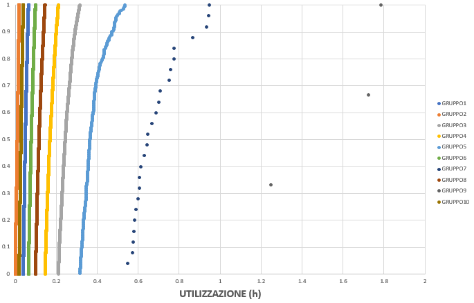
quarto d’ora

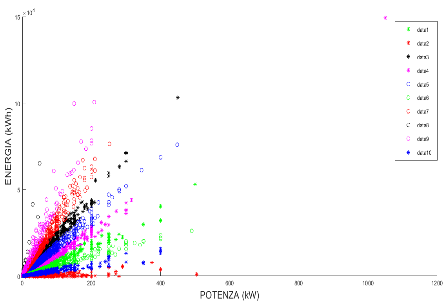
*Grafico 3: esempi di cluster creati dall’algoritmo gerarchico, macro-classe ort6, curve di carico rappresentative annuali.*

IV. CREAZIONE DEI CLUSTER UTILIZZANDO LA METODOLOGIA DI CLUSTERING K-MEANS

Un’ulteriore analisi che si è affrontata è stata quella di studiare nuovi gruppi di utenti grazie al metodo di clustering k-means, non focalizzandosi più sulle curve di carico ma studiando i diversi dati forniti per ogni utenza. Essi sono: potenza contrattuale, energia consumata mensile e l’utilizzazione. In ogni caso studio però si è variata la grandezza delle forniture sottoposta al procedimento di clustering. In un primo momento sono state create classi di utenti andando a studiare singolarmente le diverse grandezze e tramite la procedura di clustering k-means, sono stati creati cluster monodimensionali. Successivamente si è pensato di migliorare l’analisi tenendo conto di più grandezze per ogni caso studio. Questo procedimento prende il nome di clustering bidimensionale.

Per ogni caso studio analizzato sono state create differenti tipologie di gruppi sia per gli utenti trifase che monofase. I risultati possono essere ben visibili costruendo delle cumulative dei diversi cluster.





*Grafico 4: Cumulativa dell’utilizzazione, mese di aprile (dati trifase)*

*Grafico 5: Risultati del clustering bidimensionale sull’utilizzazione nel piano energia/potenza*

V. CONCLUSIONI E OBIETTIVI FUTURI

Questo lavoro è stato redatto con lo scopo di apportare un contributo allo sviluppo delle reti di distribuzione; infatti, capire l’importanza del corretto dispiegamento degli smart meter 2G sulla rete elettrica porta ad un alto numero di dati utili.

I benefici portati da questa analisi sono molteplici; conoscere le caratteristiche principali dei profili di carico delle micro-classi di clienti porta a una migliore consapevolezza sul consumo da parte dell’ente di distribuzione, che, ad esempio, può gestire al meglio la flessibilità. Inoltre, una miglior comprensione delle caratteristiche del cliente, data dalla suddivisione in classi, permetterà di stabilire un’offerta tariffaria su misura che possa aiutare i consumatori nel risparmio dei consumi.

Ulteriori sviluppi di questo studio possono portare al perfezionamento di alcune procedure sull’analisi dei dati, essenziali per rispondere alle necessità di un sistema in continua evoluzione come quello elettrico, includendo nello studio, ad esempio, carichi attivi che stanno segnando la transizione energetica permettendo una progressiva elettrificazione degli utenti finali.